Gag (Pr55^{gag}) polyprotein precursor

| | AA1 | AA133 | | | AA364 AA377 | | AA433 AA449 | AA500 |

| MA (P17) | CA (P24) | P2 | NC (P7) | P1 | P6 |
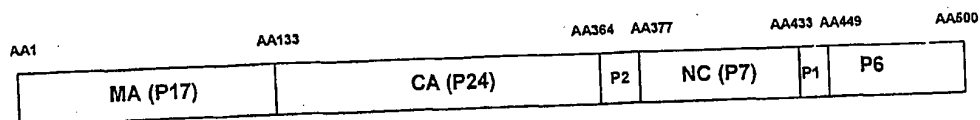
Figure 1

Matrix (MA) protein

Comprises the 3D globular domain of MA

N-myr | AA17 | AA31 | | | | | C

| Helix 1 | | Helix 2 | Helix 3 | | Helix 4 | Helix 5 |

AA16   KIRLRPGGKKKYR
Basic domain

AA99
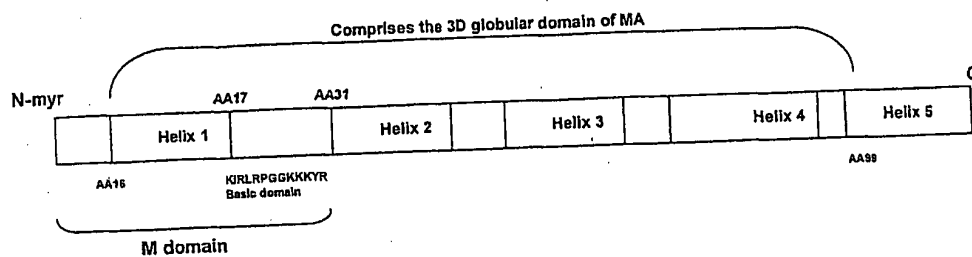
M domain

Figure 2

Nucleocapsid (NC) protein (p7)

Basic residue region   Basic residue region

N   C

AA377

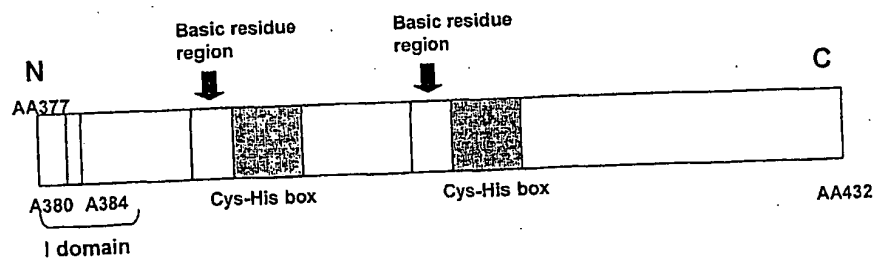A380 A384   Cys-His box   Cys-His box   AA432
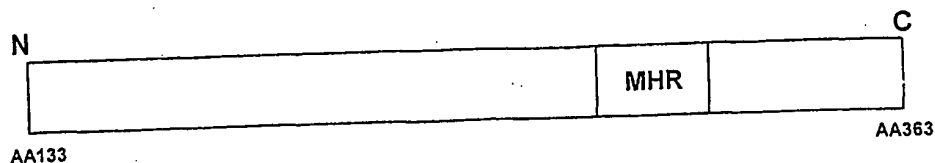
I domain

Figure 3

Capsid (CA) protein (p24)
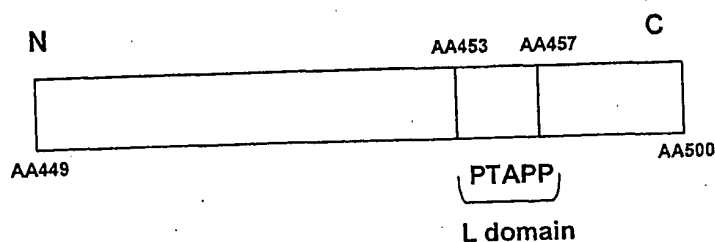

Figure 4

P6 protein

Figure 5

```
   1  GAATTCATGG GTGCGAGAGC GTCAATATTA AGAGGGGAAA AATTAGATAA ATGGGAAAAG
  61  ATTAGGTTAA GGCCAGGGGG AAAGAAACAT TATATGTTAA AACACATAGT ATGGGCGAGC
 121  AGGGAGCTGG AAAGATTTGC ACTTAACCCT GGCCTTTTAG AAACATCAGA AGGATGTAAA
 181  CAAATAATGA AACAGCTACA ACCAGCTCTC CAGACAGGAA CAGAGGAACT TAAATCATTA
 241  TACAACACAG TAGCAACTCT CTATTGTGTA CATGAAAAGA TAGAAGTACG AGACACCAAG
 301  GAAGCCTTAG ATAAGATAGA GGAAGAACAA AACAAATGTC AGCAAAAAAC GCAGCAGGCA
 361  AAAGCGGCTG ACGGGAAAGT CAGTCAAAAT TATCCTATAG TGCAGAATCT CCAAGGGCAA
 421  ATGGTACATC AAGCCATATC ACCTAGAACC TTGAATGCAT GGGTAAAAGT AATAGAAGAA
 481  AAGGCTTTTA GCCCAGAGGT AATACCCATG TTTACAGCAT TATCAGAAGG AGCCACCCCA
 541  CAAGATTTAA ACACCATGTT AAATACAGTG GGGGGACACC AAGCAGCCAT GCAAATGTTA
 601  AAAGATACTA TTAATGAAGA GGCTGCAGAA TGGGATAGAT TACATCCAGT CCATGCGGGG
 661  CCTATTGCAC CAGGCCAGAT GAGAGAACCA AGGGGAAGTG ACATAGCAGG AACTACTAGT
 721  ACCCTTCAGG AACAAATAGC ATGGATGACA AGTAACCAC CTATTCCAGT GGGAGACATC
 781  TATAAAAGAT GGATAATTCT GGGGTTAAAT AAAATAGTGA GAATGTATAG CCCGGTCAGC
 841  ATTTTGGACA TAAGACAAGG GCCAAAGGAA CCCTTTCGAG ACTATGTAGA TCGGTTCTTT
 901  AAAACTTTAA GAGCTGAACA AGCTACACAA GAAGTAAAAA ATTGGATGAC AGACACCTTG
 961  TTAGTCCAAA ATGCGAACCC AGATTGTAAG ACCATTTTGA GAGCATTAGG ACCAGGGGCT
1021  ACATTAGAAG AAATGATGAC AGCATGTCAA GGGGTGGGGA GACCTGGCCA CAAAGCAAGA
1081  GTATTGGCTG AGGCAATGAG TCAAACAAAC AGTGGAAACA TAATGATGCA GAGAAGCAAT
1141  TTTAAAGGCC CTAGAAGAAT TGTTAAATGT TTTAACTGTG GCAAGGAAGG GCACATAGCC
1201  AGAAATTGCA GAGCCCCTAG GAAAAAAGGC TGTTGGAAAT GTGGAAAAGA AGGACACCAA
1261  ATGAAAGACT GCACTGAGAG GCAGGCTAAT TTTTTAGGGA AAATTTGGCC TTCCCACAAG
1321  GGGAGGCCAG GGAATTTCCT TCAGAACAGA CCAGAGCCAA CAGCCCCACC AGCAGAGAGC
1381  TTCAGGTTCG AAGAGACAAC CCCCGCTCCG AAACAGGAGC CGATAGAAAG GGAACCCTTA
1441  ACTTCCCTCA AATCACTCTT TGGCAGCGAC CCCTTGTCTC AATAAAAGTA GGGGGCCAGA
1501  CAAGGGAGGC TCTCTTAGAC ACAGGAGCAG ATGATACAGT ATTGTCGAC
```
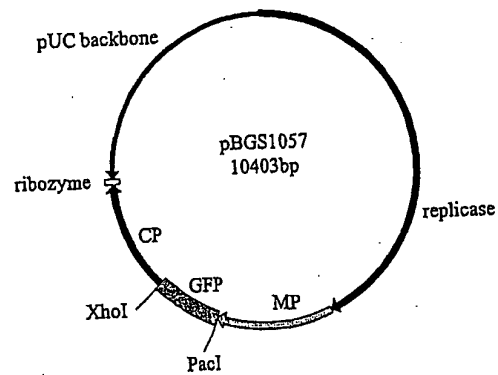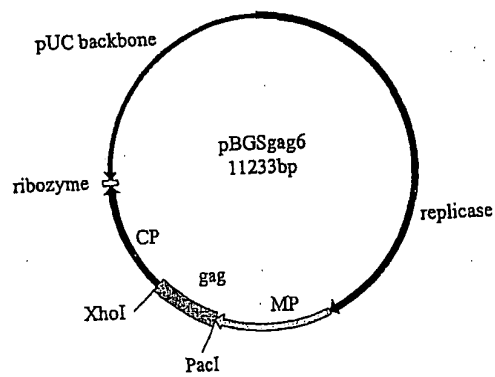
Figure 6

10/537557

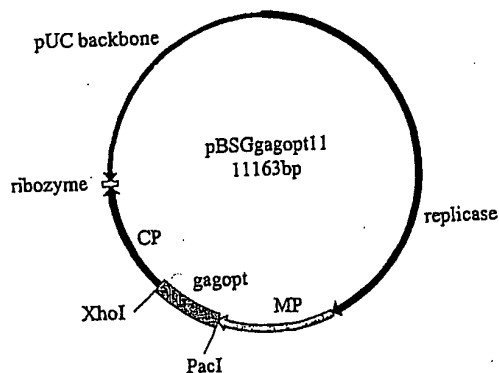3/6

Figure 7

Figure 8

10/537557

Figure 9

```
1     ATGGGTGCGA GAGCGTCAAT ATTAAGAGGG GAAAAATTAG ATAAATGGGA AAAGATTAGG
61    TTAAGGCCAG GGGGAAAGAA ACATTATATG TTAAAACACA TAGTATGGGC GAGCAGGGAG
121   CTGGAAAGAT TTGCACTTAA CCCTGGCCTT TTAGAAACAT CAGAAGGATG TAAACAAATA
181   ATGAAACAGC TACAACCAGC TCTCCAGACA GGAACAGAGG AACTTAAATC ATTATACAAC
241   ACAGTAGCAA CTCTCTATTG TGTACATGAA AAGATAGAAG TACGAGACAC CAAGGAAGCC
301   TTAGATAAGA TAGAGGAAGA ACAAAACAAA TGTCAGCAAA AAACGCAGCA GGCAAAAGCG
361   GCTGACGGGA AAGTCAGTCA AAATTATCCT ATAGTGCAGA ATCTCCAAGG GCAAATGGTA
421   CATCAAGCCA TATCACCTAG AACCTTGAAT GCATGGGTAA AAGTAATAGA AGAAAAGGCT
481   TTTAGCCCAG AGGTAATACC CATGTTTACA GCATTATCAG AAGGAGCCAC CCCACAAGAT
541   TTAAACACCA TGTTAAATAC AGTGGGGGGA CACCAAGCAG CCATGCAAAT GTTAAAAGAT
601   ACTATTAATG AAGAGGCTGC AGAATGGGAT AGATTACATC CAGTCCATGC GGGGCCTATT
661   GCACCAGGCC AGATGAGAGA ACCAAGGGGA AGTGACATAG CAGGAACTAC TAGTACCCTT
721   CAGGAACAAA TAGCATGGAT GACAAGTAAC CCACCTATTC CAGTGGGAGA CATCTATAAA
781   AGATGGATAA TTCTGGGGTT AAATAAAATA GTGAGAATGT ATAGCCCGGT CAGCATTTTG
841   GACATAAGAC AAGGGCCAAA GGAACCCTTT CGAGACTATG TAGATCGGTT CTTTAAAACT
901   TTAAGAGCTG AACAAGCTAC ACAAGAAGTA AAAAATTGGA TGACAGACAC CTTGTTAGTC
961   CAAAATGCGA ACCCAGATTG TAAGACCATT TTGAGAGCAT TAGGACCAGG GGCTACATTA
1021  GAAGAAATGA TGACAGCATG TCAGGGGGTG GGAGGACCTG GCCACAAAGC AAGAGTATTG
1081  GCTGAGGCAA TGAGTCAAAC AAACAGTGGA AACATAATGA TGCAGAGAAG CAATTTTAAA
1141  GGCCCTAGAA GAATTGTTAA ATGTTTTAAC TGTGGCAAGG AAGGGCACAT AGCCAGAAAT
1201  TGCAGAGCCC CTAGGAAAAA AGGCTGTTGG AAATGTGGAA AAGAAGGACA CCAAATGAAA
1261  GACTGCACTG AGAGGCAGGC TAATTTTTTA GGGAAAATTT GGCCTTCCCA CAAGGGGAGG
1321  CCAGGGAATT TCCTTCAGAA CAGACCAGAG CCAACAGCCC CACCAGCAGA GAGCTTCAGG
1381  TTCGAAGAGA CAACCCCCGC TCCGAAACAG GAGCCGATAG AAAGGGAACC CTTAACTTCC
1441  CTCAAATCAC TCTTTGGCAG CGACCCCTTG TCTCAATAA
```
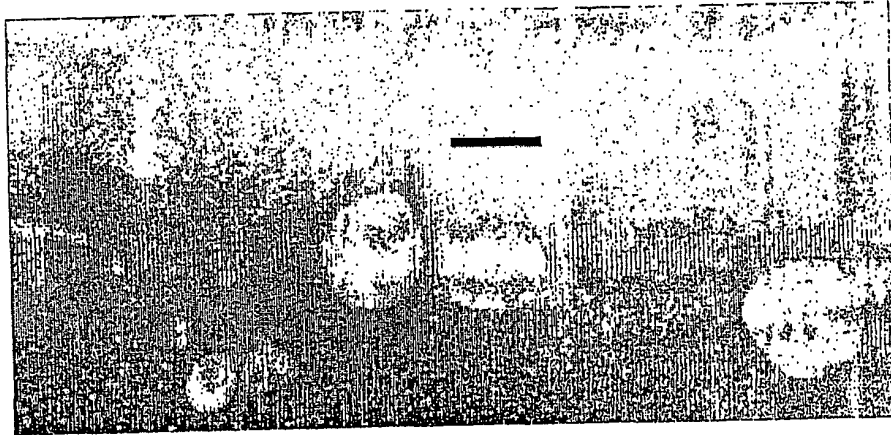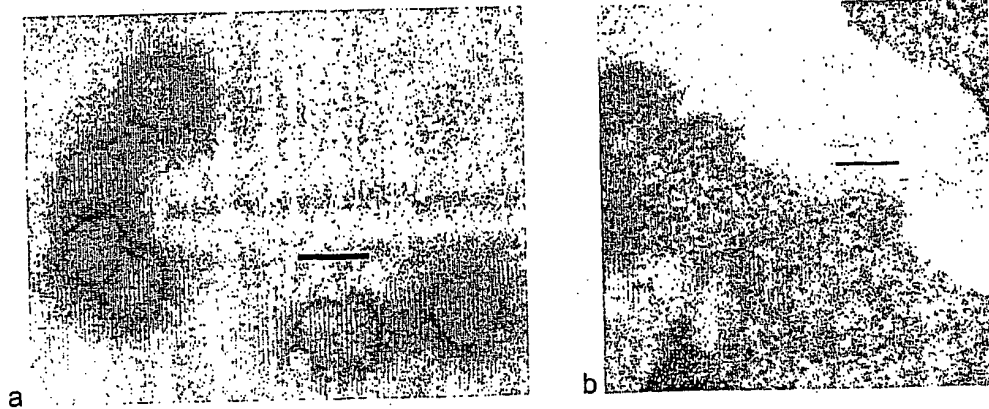
Figure 10

Figure 11

Figure 12

```
EFMGARASIL RGEKLDKWEK IRLRPGGKKH YMLKHIVWAS RELERFALNP GLLETSEGCK      60
QIMKQLQPAL QTGTEELKSL YNTVATLYCV HEKIEVRDTK EALDKIEEEQ NKCQQKTQQA     120
KAADGKVSQN YPIVQNLQGQ MVHQAISPRT LNAWVKVIEE KAFSPEVIPM FTALSEGATP     180
QDLNTMLNTV GGHQAAMQML KDTINEEAAE WDRLHPVHAG PIAPGQMREP RGSDIAGTTS     240
TLQEQIAWMT SNPPIPVGDI YKRWIILGLN KIVRMYSPVS ILDIRQGPKE PFRDYVDRFF     300
KTLRAEQATQ EVKNWMTDTL LVQNANPDCK TILRALGPGA TLEEMMTACQ GVGGPGHKAR     360
VLAEAMSQTN SGNIMMQRSN FKGPRRIVKC FNCGKEGHIA RNCRAPRKKG CWKCGKEGHQ     420
MKDCTERQAN FLGKIWPSHK GRPGNFLQNR PEPTAPPAES FRFEETTPAP KQEPIEREPL     480
TSLKSLFGSD PLSQKGARQG RLSTQEQMIQ YCR                                  513
```

Figure 13

```
MGARASILRG EKLDKWEKIR LRPGGKKHYM LKHIVWASRE LERFALNPGL LETSEGCKQI      60
MKQLQPALQT GTEELKSLYN TVATLYCVHE KIEVRDTKEA LDKIEEEQNK CQQKTQQAKA     120
ADGKVSQNYP IVQNLQGQMV HQAISPRTLN AWVKVIEEKA FSPEVIPMFT ALSEGATPQD     180
LNTMLNTVGG HQAAMQMLKD TINEEAAEWD RLHPVHAGPI APGQMREPRG SDIAGTTSTL     240
QEQIAWMTSN PPIPVGDIYK RWIILGLNKI VRMYSPVSIL DIRQGPKEPF RDYVDRFFKT     300
LRAEQATQEV KNWMTDTLLV QNANPDCKTI LRALGPGATL EEMMTACQGV GGPGHKARVL     360
AEAMSQTNSG NIMMQRSNFK GPRRIVKCFN CGKEGHIARN CRAPRKKGCW KCGKEGHQMK     420
DCTERQANFL GKIWPSHKGR PGNFLQNRPE PTAPPAESFR FEETTPAPKQ EPIEREPLTS     480
LKSLFGSDPL SQ                                                        
```

Figure 14